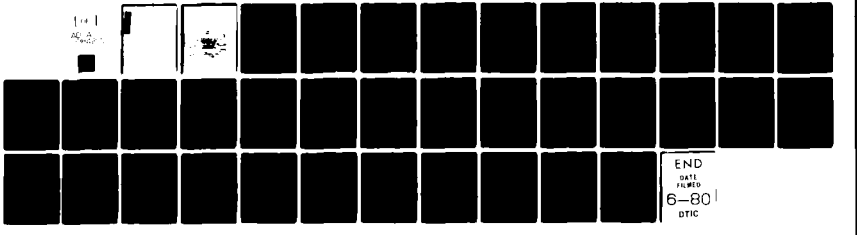
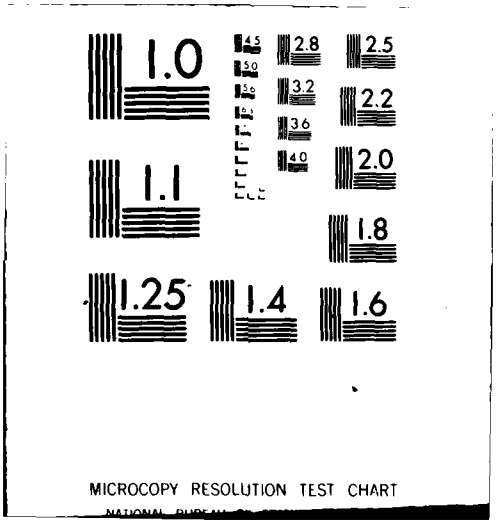
AD-A084 272 MASSACHUSETTS INST OF TECH LEXINGTON LINCOLN LAB F/G 17/7
APPLICATION OF LINEAR PREDICTION TO MULTICLASS SEQUENTIAL DISCR--ETC(U)
DEC 79 C W THERRIEN F19628-80-C-0002
UNCLASSIFIED TN-1979-63 ESD-TR-79-308 NL



END
DATE FILMED
6-80
DTIC


1.0
1.1
1.25
1.4
1.6
1.8
2.0
2.2
2.5
2.8
3.2
3.6
4.0
4.5
5.0
5.6
MICROCOPY RESOLUTION TEST CHART

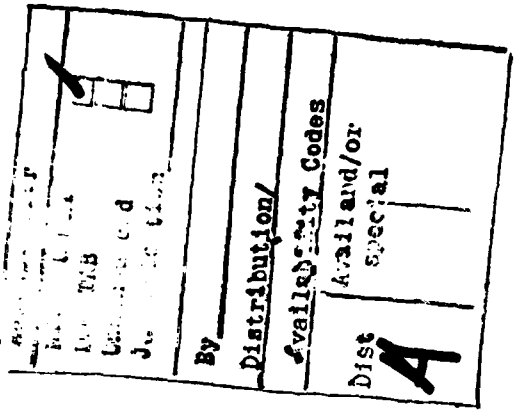
MASSACHUSETTS INSTITUTE OF TECHNOLOGY

LINCOLN LABORATORY

APPLICATION OF LINEAR PREDICTION TO MULTICLASS SEQUENTIAL DISCRIMINATION

C. W. THERRIEN

Group 92



TECHNICAL NOTE 1979-63

19 DECEMBER 1979

LEXINGTON MASSACHUSETTS

ABSTRACT

A linear predictive form of classifier is described for the sequential discrimination of target signatures. The classifier implements a Generalized Sequential Probability Ratio Test (GSPRT) which allows for multiple classes of target signatures. Forms of the classifier for both noncoherent and coherent signatures are described. Flow charts are included for convenience of implementation.

CONTENTS

I. INTRODUCTION

An earlier report [1] and an associated paper [2] described a linear predictive form of a two-class sequential classifier for target signature data. The present report describes extensions of that sequential classifier to multiple classes and to coherent signature data. The new results, although complete in themselves, draw heavily on material presented in the other two publications. This report, therefore, assumes that the reader is familiar with the material in References 1 and 2.

The extensions described here are conceptually straightforward and have already been cited in the previous publications. However, results are presented here in sufficient detail for immediate implementation. Section II describes the multiclass sequential classifier for noncoherent signatures and Section III formulates the results for coherent signatures. Section IV is a summary. In addition, two appendixes are included. Appendix A provides background material on complex Gaussian processes that justifies the coherent formulation of the classifier. Appendix B suggests a further extension of the classifier to a class of nonstationary signatures.

II. MULTICLASS SEQUENTIAL CLASSIFIER

A. Statistical Theory

Multiclass sequential testing - The multiclass sequential classifier is based on the generalized sequential probability ratio test (GSPRT) due to Reed [3] and represents an extension of the Wald sequential probability ratio test (SPRT) for two hypotheses [4]. In the context of target classification let $\underline{X}_k$ represent the vector consisting of a set of k consecutive returns $\underline{x}_1, \underline{x}_2, \ldots, \underline{x}_k$ on a target which are equally spaced in time. Such a set of returns is called a target signature. The individual returns $\underline{x}_j$ are in general vector quantities since the sensor may measure several quantities at once (radars may be dual polarized and samples may be taken in multiple gates, optical sensors may take measurements in several spectral bands, etc.). Let $p_i(\underline{X}_k)$ represent the probability density function under the hypothesis H_i that a target belongs to class i. Then the GSPRT proceeds by successive rejection of hypotheses as follows.

Let C be the total number of hypotheses (classes) under consideration and let C_k be the number of hypotheses that have not been rejected before k returns. Then a set of C_k generalized likelihood ratios is defined by

$$\lambda_k^i(\underline{X}_k) = p_i(\underline{X}_k) \Big/ \Big[\prod_{\substack{j \text{ not} \\ \text{rejected}}} p_j(\underline{X}_k)\Big]^{\frac{1}{C_k}}$$

$$\forall i \ni H_i \epsilon \{\text{hypotheses not rejected}\} \qquad (2.1)$$

and a corresponding set of thresholds A_k^i is selected in a manner to be discussed presently. The following set of tests is then performed

$$\lambda_k^i(\underline{X}_k) < A_k^i \longrightarrow \text{reject } H_i \qquad (2.2)$$

That is, all of the hypotheses are tested and a given hypothesis H_i is rejected if the corresponding generalized likelihood ratio is less than its threshold. If this procedure results in only one unrejected hypothesis, that hypothesis is the one accepted. Otherwise another return $\underline{x}_{k+1}$ is observed and the procedure is repeated until there is only a single hypothesis remaining.

Reed suggests that the thresholds for the multi-class sequential test be computed from the relation

$$A_k^i = \Big[\sum_{\substack{\ell \text{ not} \\ \text{rejected} \\ \ell \neq i}} e_{\ell i}\Big] \Big/ \Big[\prod_{j \text{ not rejected}} (1 - e_{ij})\Big]^{\frac{1}{C_k}}$$

$$i \ni H_i \epsilon \{\text{hypotheses not rejected}\} \qquad (2.3)$$

where e_{ij} is the (desired) probability of deciding upon

H_i given that H_j is correct. Observe that this requires recomputation of the thresholds after each rejection of a set of hypotheses. Since, except in the two-class case, the GSPRT does not have any known theoretically optimal properties, in practice it is just as reasonable to determine the initial thresholds from (2.3) and keep the thresholds fixed or to allow the thresholds to decrease in a prescribed manner with the number of observations. A convenient family of these time-varying thresholds is given by [5]

$$A_k^i = \exp\left[-\frac{1}{2} T_o^i \left(1-\frac{k-1}{k_{max}-1}\right)^{\eta_i}\right], \quad 0<\eta_i<1 \tag{2.4}$$

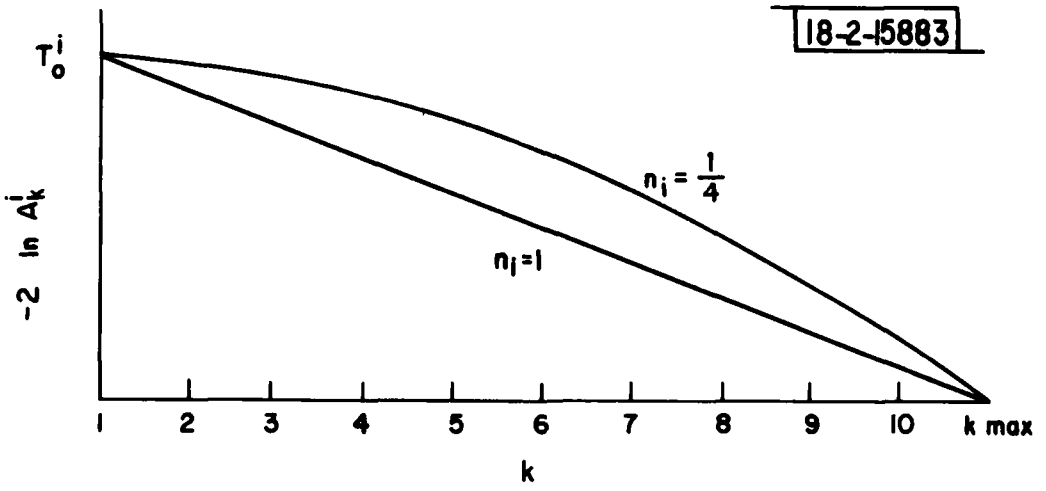


Fig. 2.1 - Form of Time-Varying Thresholds

and is plotted in Fig. 2.1 for two values of η_i. Time-varying thresholds tend to decrease the average number of returns necessary to make a classification, but otherwise have no specific advantages.

Observe that neither the fixed nor the time-varying thresholds guarantee that a decision will be made after a finite number of observations. Thus in practical applications it is necessary to truncate the test after a fixed number of observations. A decision can then be conveniently made by the principle of maximum likelihood. Hypothesis H_i is chosen where

$$p_i(\underline{X}_k)_{max} = \max_{\substack{j \text{ not} \\ \text{rejected}}} \left\{ p_j(\underline{X}_k)_{max} \right\} \tag{2.5}$$

Alternatively a Bayes or Neyman-Pearson test can be formulated involving the quantities $p_i(\underline{X}_k)_{max}$, and that test can be evaluated to make the classification decision. Details can be found in various texts in engineering and statistics (see, e.g., Refs. 6 and 7).

Relations to two-class sequential testing - As stated earlier, the GSPRT is an extention of the Wald SPRT. The SPRT is conventionally stated as

$$\ell_k(\underline{X}_k) = \frac{p_1(\underline{X}_k)}{p_2(\underline{X}_k)} \quad \begin{cases} > A \longrightarrow \text{accept } H_1 \\ < B \longrightarrow \text{accept } H_2 \end{cases} \tag{2.6}$$

where

$$A = \frac{1-e_{21}}{e_{12}}$$
$$B = \frac{e_{21}}{1-e_{12}} \tag{2.7}$$

For tests involving two hypotheses (C=2), the GSPRT and the SPRT are equivalent and the quantities defining them are related by

$$\ell_k(\underline{X}_k) = \left[\lambda_k^1(\underline{X}_k)\right]^2 = \left[1/\lambda_k^2(\underline{X}_k)\right]^2 \tag{2.8}$$

and

$$A = \left[\frac{1}{A^2}\right]^2$$
$$B = \left[A^1\right]^2 \tag{2.9}$$

In this case both the SPRT and the GSPRT satisfy an optimality condition. In particular they minimize the average number of returns necessary to achieve the specified error probabilities e_{12} and e_{21}.

B. Linear Predictive Formulation

It is easy to show (see Ref. 1) that if the observations $\underline{x}_1, \underline{x}_2, \ldots, \underline{x}_k$ are zero mean and jointly Gaussian then the log density function can be written in the recursive form

$$\ell n\ p(\underline{X}_k) = \ell n\ p(\underline{X}_{k-1}) - \frac{1}{2}\underline{\varepsilon}_k^T E_k^{-1} \underline{\varepsilon}_k - \frac{1}{2}\ell n\ |E_k| - \frac{1}{2} n \ell n 2\pi \tag{2.10}$$

where

$$\underline{\varepsilon}_k = \underline{x}_k - G_k^T \underline{X}_{k-1} = \left[-G_k^T \;\vdots\; I\right]\underline{X}_k \tag{2.11}$$

and G_k and E_k are matrix parameters derived from the covariance matrix of $\underline{X}_k$ and n is the dimension of the $\underline{x}_i$. The quantity $\hat{\underline{x}}_k = G_k^T \underline{X}_{k-1}$ is the optimal linear estimate of $\underline{x}_k$ given $\underline{X}_{k-1}$, and G_k is the matrix of prediction coefficients that produces that estimate. The quantity $\underline{\varepsilon}_k$ is the error $\underline{x}_k - \hat{\underline{x}}_k$ in the estimate and E_k is its covariance matrix. Both G_k and E_k are computed from training data in a manner to be described later. If the observations are not of zero mean then a simple removal of the mean of $\underline{x}_k$ and $\underline{X}_{k-1}$ in (2.11) allows for the representation in (2.10). Under conditions of stationarity the classifier matrix parameters G_k and E_k for large k can be well approximated by matrices of lower order p+1 and this greatly simplifies the classifier. For values of k greater than p+1, only the last p returns $\underline{x}_{k-p}, \ldots \underline{x}_{k-1}$ are used to form the estimate $\hat{\underline{x}}_k$. The selected value p

relates to the modeling of target signatures as a multivariate autoregressive process and is beyond the scope of the present discussion. However Refs. [8] and [9] give approaches to and discussions of the modeling problem.

Equations (2.1), (2.2), (2.10), and (2.11) specify the classification algorithm. The classifier is represented as a structured flow chart in Table 2.1 and as a block diagram in Fig. 2.2. In the figure, (2.11) is realized by multichannel discrete-time filters. The filters are time-varying for $k \leq p+1$. However, for $k > p+1$ the filters are time-invariant and have impulse response of finite duration $p+1$.

The multiclass sequential classifier has a simple intuitive explanation which follows from Fig. 2.2 and the equations in Table 2.1. A signature is applied to all branches of the classifier simultaneously. In each branch, the mean of the target class is removed, and the signature is fed through a linear filter which produces the prediction error process $\underline{\varepsilon}_k$. The terms $\underline{\varepsilon}_k^T E^{-1} \underline{\varepsilon}_k$ appearing in Table 2.1, when summed over k, represent a normalized energy in the error process. When this sum (the integrated error) for a given branch exceeds the average of the corresponding sums for all branches by a

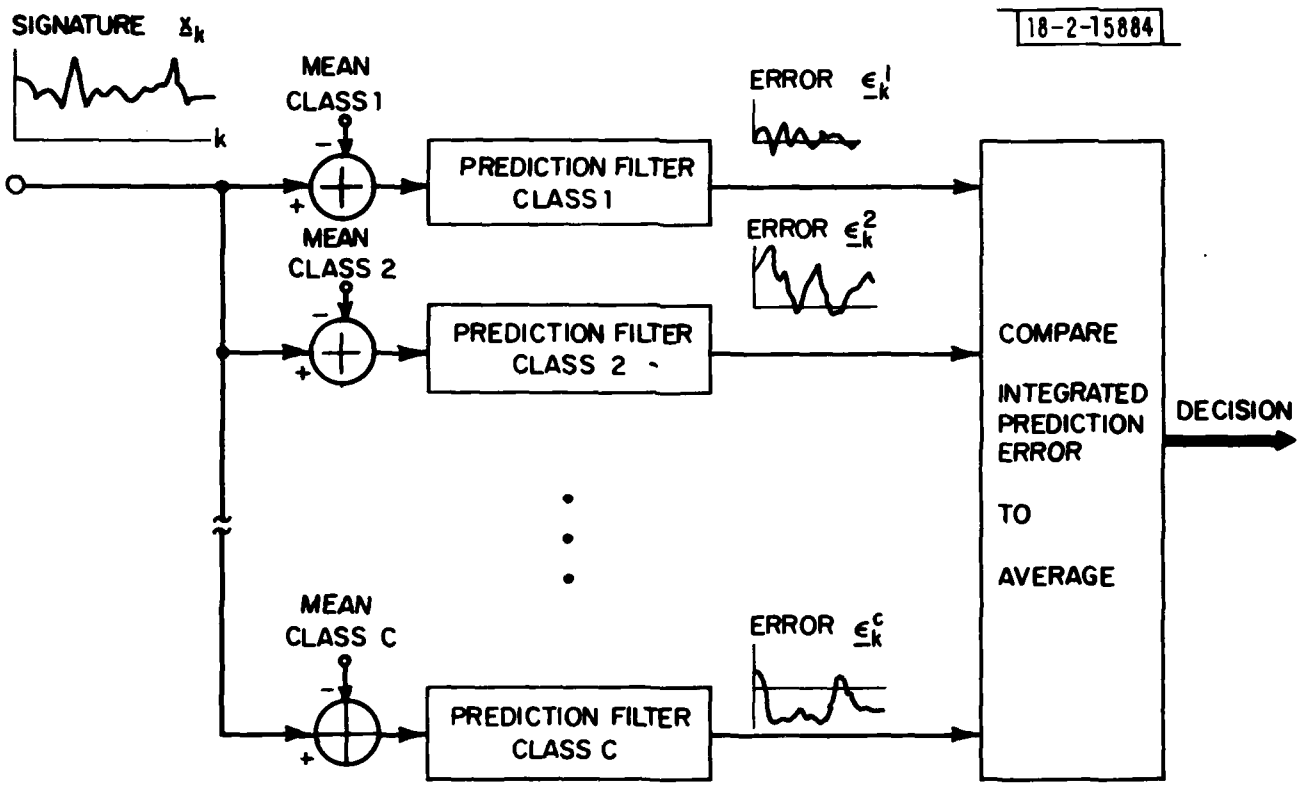
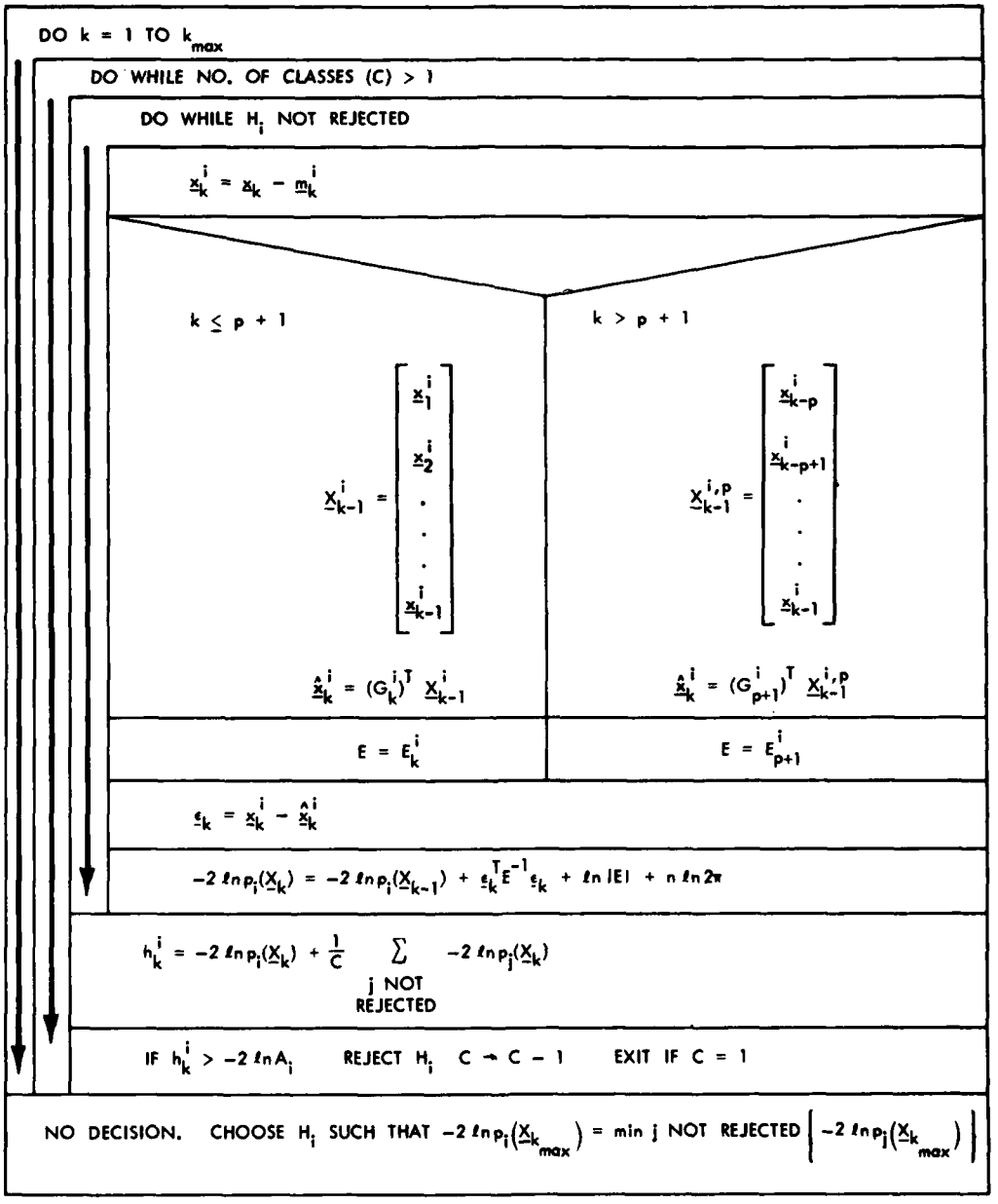


Fig. 2.2. Multiclass sequential classifier.

TABLE 2.1

STRUCTURED FLOW CHART FOR MULTICLASS SEQUENTIAL CLASSIFIER



10-2-15886

predetermined threshold* the class represented by the branch is rejected. This procedure continues until a single class remains unrejected.

The classifier parameters G_k, E_k, k = 1, 2, ..., p+1, are most efficiently computed from the matrix covariance function $R(\ell) = \mathcal{E}\left[\underline{x}_k \underline{x}_{k-\ell}^T\right]$ via the multivariate Levinson recursion [10] or directly from training data using various algorithms (see, e.g., Ref., 11).** The formulas involve an auxiliary matrix parameter Γ (the reflection coefficient) and a set of backward prediction parameters $G_k^\dagger$, $E_k^\dagger$, $\Gamma_k^\dagger$ which pertain to the problem of estimating $\underline{x}_1$ from $\underline{x}_2$, ..., $\underline{x}_k$. Although these additional parameters are not used in the classifier, they must nevertheless be carried along in the recursion for the classifier parameters. The formulas that define the multivariate Levinson recursion are as follows.

---

*For purposes of this intuitive explanation it is assumed that the terms $\ell n|E|$ and $n\ell n\, 2\Pi$ are lumped into the threshold.

**Both methods assume stationarity. A less efficient method that does not assume stationarity is given in Ref. 1.

$$G_k = \begin{bmatrix} 0 \\ \hline G_{k-1} \end{bmatrix} + \begin{bmatrix} I \\ \hline -G_{k-1}^{\dagger} \end{bmatrix} \Gamma_{k-1}$$

$$G_k^{\dagger} = \begin{bmatrix} G_{k-1}^{\dagger} \\ \hline 0 \end{bmatrix} + \begin{bmatrix} -G_{k-1} \\ \hline I \end{bmatrix} \Gamma_{k-1}^{\dagger} \qquad (2.12)$$

where

$$\Gamma_k = (E_k^{\dagger -1})(R^T(k) - G_k^{\dagger T} B_k)$$

$$\Gamma_k^{\dagger} = E_k^{-1}(R^T(k) - G_k^{\dagger T} B_k)^T \qquad (2.13)$$

and

$$E_k = E_{k-1} - \Gamma_{k-1}^T E_{k-1}^{\dagger} \Gamma_{k-1}$$

$$E_k^{\dagger} = E_{k-1}^{\dagger} - \Gamma_{k-1}^{\dagger T} E_{k-1} \Gamma_{k-1}^{\dagger} \qquad (2.14)$$

When the signature returns $\underline{x}_j$ are one-dimensional (n=1) the forward and backward prediction parameters are identical and each pair of equations (2.12) - (2.14) reduces to a single equation. The resulting three equations define the usual (univariate) Levinson recursion.

III. MULTICLASS SEQUENTIAL CLASSIFIER FOR COHERENT SIGNATURES

A. Motivation for the Coherent Form

The signature measured by a coherent radar can be represented by a complex random process. Since the classifier developed in Section II allows for multidimensional observations, the real and imaginary parts of the complex process can be represented by two correlated real processes and a classifier can be developed on that basis. However, the complex process that represents a radar signature is the complex envelope of a real-valued narrowband process at the video stage of a radar receiver. If this signal is stationary then the usual theory for complex processes can be applied to the signature. This allows for a more simple and elegant analysis than is possible by treating the components of the signature as correlated real processes. The relations between the two approaches are discussed briefly in Appendix A. The topic is discussed in more detail in several references [12-14].

In this section the individual coherent measurements are represented by vector quantities $\underline{z}_k = \underline{x}_k + j\ \underline{y}_k$ where $\underline{x}_k$ and $\underline{y}_k$ are the quadrature components of the signal. We hasten to point out that $\underline{x}_k$ as used here does not represent the same quantity used in Section II and that the classifier for real

processes cannot be considered to be a special case of the classifier for coherent processes.

B. Coherent Form of the Classifier

Following the development in Section II, let $\underline{Z}_k$ represent the set of observations $\underline{z}_1, \underline{z}_2, \ldots, \underline{z}_k$. Then (see Appendix A) the joint Gaussian density function for the pairs $\underline{x}_1, \underline{y}_1, \underline{x}_2, \underline{y}_2, \ldots, \underline{x}_k, \underline{y}_k$ is completely equivalent to the complex Gaussian density function

$$p(\underline{Z}_k) = \frac{1}{\pi^{nk}\,|K|} e^{-(\underline{Z}_k-\bar{\underline{Z}}_k)^* K^{-1}(\underline{Z}_k-\bar{\underline{Z}}_k)} \tag{3.1}$$

where $\bar{\underline{Z}} = \mathcal{E}[\underline{Z}]$ and K is the covariance matrix $\mathcal{E}[(\underline{Z}-\bar{\underline{Z}})(\underline{Z}-\bar{\underline{Z}})^*]$. Although in most applications the complex random vector will have zero mean (see discussion in Appendix A) we include the mean here for the sake of generality. Then paralleling the development in Ref. 1 it can be shown that

$$\ell n\, p(\underline{Z}_k) = \ell n\, p(\underline{Z}_{k-1}) - \underline{\varepsilon}_k^* E_k^{-1} \underline{\varepsilon}_k - \ell n\, |E_k| - n\, \ell n\, \pi \tag{3.2}$$

where

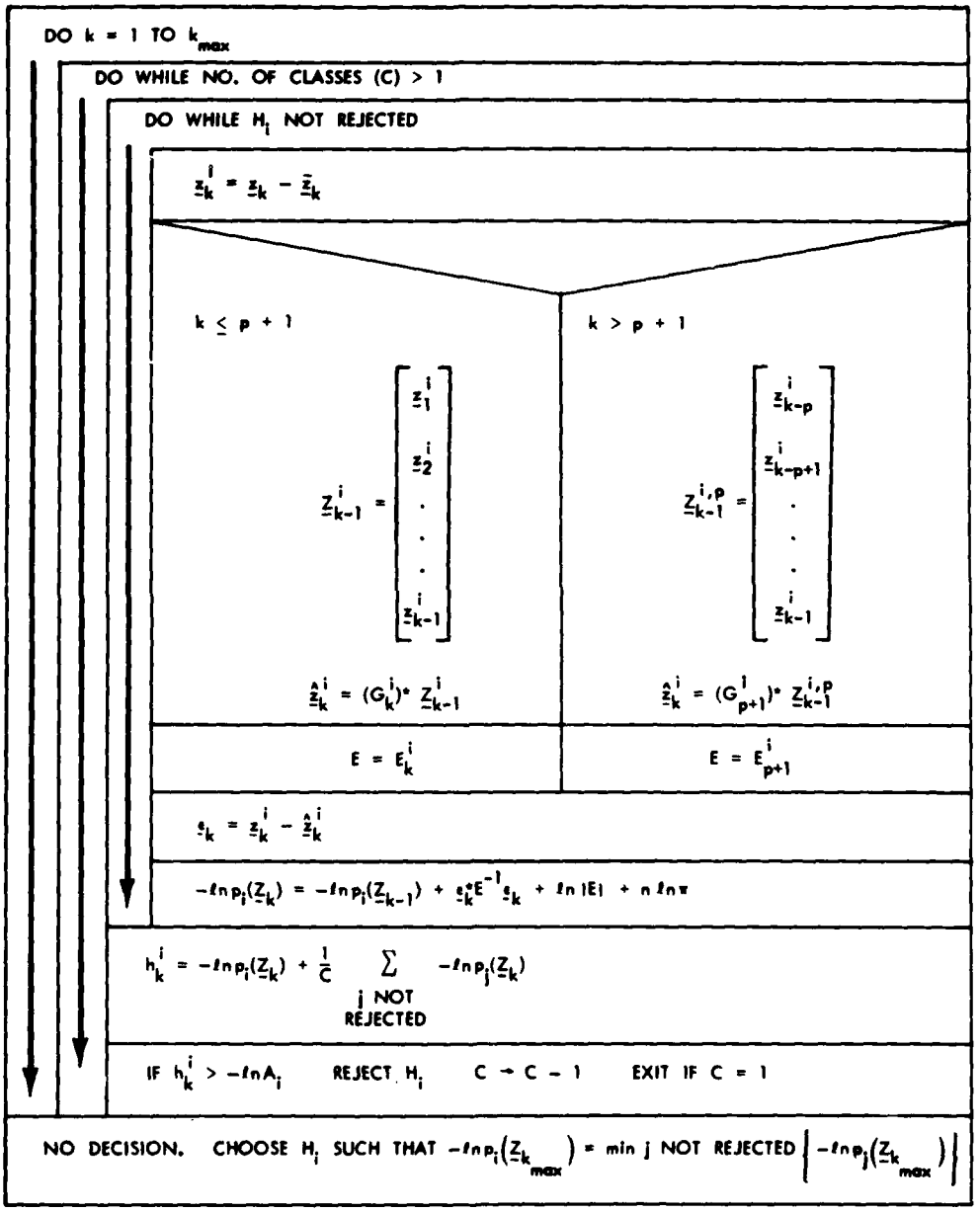
$$\underline{\varepsilon}_k = \underline{z}_k' - G_k^* \underline{Z}_{k-1}' = [-G_k^* \,\vdots\, I]\,\underline{Z}_k' \tag{3.3}$$

where $\underline{z}'_k$ and $\underline{Z}'_k$ represent observations after removal of the mean, and G_k and E_k are derived from the covariance matrix of the observations in the same manner as they are for the noncoherent case. The flow chart for the coherent form of the classifier is given in Table 3.1. Figure 2.1 and the interpretation given in Section II still applies although in this case the filters have complex coefficients.

The classifier parameters can most conveniently be computed from the complex form of the multivariate Levinson recursion. Since the relevant equations are obtained from (2.12) - (2.14) by merely replacing the transpose of a matrix by its transpose conjugate, those equations need not be repeated here. As in the noncoherent case, if the signature returns $\underline{z}_j$ are one-dimensional the second equation in each pair of (2.12) - (2.14) is redundant and can be eliminated.

TABLE 3.1

STRUCTURED FLOW CHART FOR THE COHERENT FORM OF THE MULTICLASS SEQUENTIAL CLASSIFIER



18-2-15887

IV. SUMMARY

This report described a multiclass sequential classifier for signatures of targets observed by a radar or other sensor. The classifier implementation was based on linear predictive filtering concepts. A two-class version of this classifier had been developed earlier. The current report extends the work to multiple classes and to coherent signature data.

APPENDIX A

COMPLEX REPRESENTATION FOR STATIONARY COHERENT PROCESSES

Define an n-dimensional complex vector random process $\underline{z}(t) = \underline{x}(t)+j\underline{y}(t)$ where $\underline{x}(t)$ and $\underline{y}(t)$ are n-dimensional real jointly Gaussian vector random processes. Let $\underline{Z}$,$\underline{X}$, and $\underline{Y}$ represent the nk-dimensional vectors formed from k samples of the processes. We have

$$\underline{Z} = \underline{X} + j\underline{Y} \tag{A.1}$$

Define also

$$\underline{\xi} = \begin{bmatrix} \underline{X} \\ \underline{Y} \end{bmatrix} \tag{A.2}$$

so that

$$\underline{Z} = [I \,|\, jI]\ \underline{\xi} \tag{A.3}$$

Then it follows that the covariance matrices for $\underline{\xi}$ and $\underline{Z}$ are respectively

$$\Sigma = \mathcal{E}\left[(\underline{\xi}-\bar{\underline{\xi}})(\underline{\xi}-\bar{\underline{\xi}})^T\right] = \left[\begin{array}{c|c} \Sigma_x & \Sigma_{xy} \\ \hline \Sigma_{yx} & \Sigma_y \end{array}\right] \tag{A.4}$$

$$K = \mathcal{E}\left[(\underline{Z}-\bar{\underline{Z}})(\underline{Z}-\bar{\underline{Z}})^*\right] = [I \,|\, jI] \left[\begin{array}{c|c} \Sigma_x & \Sigma_{xy} \\ \hline \Sigma_{yx} & \Sigma_y \end{array}\right] \left[\begin{array}{c} I \\ \hline -jI \end{array}\right]$$

$$= (\Sigma_x + \Sigma_y) + j(\Sigma_{yx} - \Sigma_{xy}) \tag{A.5}$$

where the bar over the variables represents the mean value. If the processes $\underline{x}(t)$ and $\underline{y}(t)$ are the quadrature components of a narrowband process

$$\underline{v}(t) = \underline{x}(t) \cos \omega_o t + \underline{y}(t) \sin \omega_o t \tag{A.6}$$

as in the case of a radar signal then $\underline{z}(t)$ represents the complex envelope of the process. If the process $\underline{v}(t)$ is stationary, then the complex envelope $\underline{z}(t)$ (i.e., both $\underline{x}(t)$ and $\underline{y}(t)$) must have zero mean. Further, it is easy to show by expanding $\mathcal{E}[\underline{v}(t+\tau)\ \underline{v}^T(t)]$ (see e.g., Ref. 12) that if $\underline{v}(t)$ is stationary then $\underline{x}(t)$ and $\underline{y}(t)$ must be jointly stationary and their covariance functions must satisfy symmetry conditions $K_x(\tau)=K_y(\tau)$ and $K_{xy}(\tau)=-K_{yx}(\tau)$. It therefore follows that

$$\begin{aligned} \Sigma_x &= \Sigma_y \\ \Sigma_{xy} &= -\Sigma_{yx} \end{aligned} \tag{A.7}$$

and thus (from (A.5)) that

$$K = 2\Sigma_x + j2\Sigma_{yx} \tag{A.8}$$

Under these conditions the Hermitian covariance matrix K provides a complete description of the vectors $\underline{X}$ and $\underline{Y}$ with respect to their individual second moment characteristics and with respect to their cross correlations. Further, the analysis is simpler and more elegant in terms of the

complex representation and hypothesis testing such as that performed by the sequential classifier can be more efficiently carried out by treating $\underline{x}(t)$ and $\underline{y}(t)$ as components of a stationary complex process.

Since in many practical problems the complex envelope $\underline{z}(t)$ arising from a radar signal will have random phase, the zero mean condition is not a severe restriction. If $\underline{z}(t)$ does not have random phase the process $\underline{v}(t)$ can be made stationary (provided that conditions (A.7) hold) by removing the time-varying function

$$\bar{\underline{x}} \cos \omega_0 t + \bar{\underline{y}} \sin \omega_0 t \tag{A.9}$$

This implies that a phase coherency is maintained between the transmitted and received signals in such a way that $\bar{\underline{x}}$ and $\bar{\underline{y}}$ can be properly associated with the quadrature components. Although this may be difficult to achieve in practice, we shall allow for a non-zero mean of the complex vector $\underline{Z}$ in order to maintain generality.

It is the goal of the following to show that if conditions (A.7) hold then the multivariate Gaussian density function

$$P_{\underline{\xi}}(\underline{\xi}) = \frac{1}{(2\pi)^{\frac{2nk}{2}} |\Sigma|^{\frac{1}{2}}} e^{-\frac{1}{2}(\underline{\xi}-\bar{\underline{\xi}})^T \Sigma^{-1} (\underline{\xi}-\bar{\underline{\xi}})} \tag{A.10}$$

is completely equivalent to the complex Gaussian density function

$$p_{\underline{Z}}(\underline{Z}) = \frac{1}{\pi^{nk}|K|} e^{-(\underline{Z}-\bar{\underline{Z}})^* K^{-1}(\underline{Z}-\bar{\underline{Z}})} \tag{A.11}$$

in the sense that is if $\underline{Z}_o$ corresponds to $\underline{\xi}_o$ then $p_{\underline{Z}}(\underline{Z}_o) = p_{\underline{\xi}}(\underline{\xi}_o)$. This will be done in two steps.

First observe that from (A.4) and (A.8) we can write

$$\sum = \frac{1}{2}\mathrm{Re}\begin{bmatrix} K & jK \\ -jK & K \end{bmatrix} = \frac{1}{2}\mathrm{Re}\left\{ \begin{bmatrix} I \\ -jI \end{bmatrix} K \begin{bmatrix} I & jI \end{bmatrix} \right\} \tag{A.12}$$

By using (A.12) it can be directly verified that

$$\sum^{-1} = 2\ \mathrm{Re}\left\{ \begin{bmatrix} I \\ -jI \end{bmatrix} K^{-1} \begin{bmatrix} I & jI \end{bmatrix} \right\} \tag{A.13}$$

Thus the exponent in (A.10) is

$$-\frac{1}{2}(\underline{\xi}-\bar{\underline{\xi}})\sum^{-1}(\underline{\xi}-\bar{\underline{\xi}}) = -\frac{1}{2}2\mathrm{Re}\left[\begin{bmatrix} (\underline{X}-\bar{\underline{X}})^T & (\underline{Y}-\bar{\underline{Y}})^T \end{bmatrix} \begin{bmatrix} I \\ -jI \end{bmatrix} K^{-1} \begin{bmatrix} I & jI \end{bmatrix} \begin{bmatrix} \underline{X}-\bar{\underline{X}} \\ \underline{Y}-\bar{\underline{Y}} \end{bmatrix} \right]$$

$$= -\mathrm{Re}\ (\underline{Z}-\bar{\underline{Z}})^* K^{-1}(\underline{Z}-\bar{\underline{Z}}) = -(\underline{Z}-\bar{\underline{Z}})^* K^{-1}(\underline{Z}-\bar{\underline{Z}}) \tag{A.14}$$

(The last equality follows from the fact that K is Hermitian.)

Secondly, it can be shown that

$$|\sum| = \left(\frac{1}{2}\right)^{2nk}|K|^2 \tag{A.15}$$

which implies that the scale factors in the two density functions are equal. In order to show this, observe that the determinant of K can be expressed as

$$|K| = \prod_{i=1}^{nk} \lambda_i \tag{A.16}$$

where λ_i are the eigenvalues of K, which are real but not necessarily distinct. Now let $\underline{e}$ be an eigenvector of K corresponding to an eigenvalue λ. We shall show that the vectors

$$\underline{\eta}_R = \mathrm{Re}\begin{bmatrix} \underline{e} \\ -j\underline{e} \end{bmatrix}$$

and (A.17)

$$\underline{\eta}_I = \mathrm{Im}\begin{bmatrix} \underline{e} \\ -j\underline{e} \end{bmatrix}$$

are eigenvectors of Σ both corresponding to an eigenvalue $\lambda/2$. (It is trivial to show that $\underline{\eta}_R$ and $\underline{\eta}_I$ are orthogonal. Since eigenvectors $\underline{e}_i$ corresponding to various λ_i are orthogonal, a straightforward computation shows that the $\underline{\eta}_R$ and $\underline{\eta}_I$ corresponding to different eigenvalues are also mutually orthogonal and thus constitute a complete set of eigenvectors for Σ.)

That (A.17) are eigenvectors can be shown as follows. From (A.12) and (A.17) we have

$$\Sigma \underline{\eta}_R = \frac{1}{2} \left[\begin{array}{c|c} \text{Re } K & -\text{Im } K \\ \hline \text{Im } K & \text{Re } K \end{array}\right] \cdot \begin{bmatrix} \text{Re } \underline{e} \\ \text{Im } \underline{e} \end{bmatrix}$$

$$= \frac{1}{2} \begin{bmatrix} \text{Re } K \text{ Re } \underline{e} - \text{Im } K \text{ Im } \underline{e} \\ \text{Im } K \text{ Re } \underline{e} + \text{Re } K \text{ Im } \underline{e} \end{bmatrix}$$

$$= \frac{1}{2} \begin{bmatrix} \text{Re } K\underline{e} \\ \text{Im } K\underline{e} \end{bmatrix} = \frac{\lambda}{2} \begin{bmatrix} \text{Re } \underline{e} \\ \text{Im } \underline{e} \end{bmatrix} = \frac{\lambda}{2} \underline{\eta}_R$$

and

$$\Sigma \underline{\eta}_I = \frac{1}{2} \left[\begin{array}{c|c} \text{Re } K & -\text{Im } K \\ \hline \text{Im } K & \text{Re } K \end{array}\right] \cdot \begin{bmatrix} \text{Im } \underline{e} \\ -\text{Re } \underline{e} \end{bmatrix}$$

$$= \frac{1}{2} \begin{bmatrix} \text{Re } K \text{ Im } \underline{e} + \text{Im } K \text{ Re } \underline{e} \\ \text{Im } K \text{ Im } \underline{e} - \text{Re } K \text{ Re } \underline{e} \end{bmatrix}$$

$$= \frac{1}{2} \begin{bmatrix} \text{Im } K\underline{e} \\ -\text{Re } K\underline{e} \end{bmatrix} = \frac{\lambda}{2} \begin{bmatrix} \text{Im } \underline{e} \\ -\text{Re } \underline{e} \end{bmatrix} = \frac{\lambda}{2} \underline{\eta}_I$$

Since Σ has two identical eigenvalues for every eigenvalue of K, the determinant of Σ is given by

$$|\Sigma| = \prod_{i=1}^{nk} \left(\frac{\lambda_i}{2}\right)^2 = \left(\frac{1}{2}\right)^{2nk} |K|^2 \qquad \text{(A.18)}$$

Thus the equivalence of (A.10) and (A.11) is shown.

APPENDIX B

EXTENSION OF THE SEQUENTIAL CLASSIFIER TO A CLASS OF NONSTATIONARY SIGNATURES

Let $\underline{\boldsymbol{\epsilon}}_k$ represent the series of terms $\underline{\varepsilon}_1, \underline{\varepsilon}_2, \ldots, \underline{\varepsilon}_k$ defined in (2.11). It is obvious from (2.11) that the quantities $\underline{\boldsymbol{\epsilon}}_k$ and $\underline{X}_k$ are related by a causal linear transformation with Jacobian equal to one. Thus it is possible to write

$$p_{\underline{X}}(\underline{X}_k) = p_{\underline{\boldsymbol{\epsilon}}}(\underline{\boldsymbol{\epsilon}}_k) \tag{B.1}$$

for any $\underline{\boldsymbol{\epsilon}}_k$ corresponding to a given $\underline{X}_k$. The fact that the $\{\underline{\varepsilon}_k\}$ are uncorrelated (see Ref. 1 or 2) allows a simple factorization of the density $p_{\underline{\boldsymbol{\epsilon}}}$ and this leads directly to the representation in (2.10) and the associated classifier structure.

While strictly speaking the classifier developed in Section II permits classification of nonstationary signatures, many of the advantages of the classifier are lost. The filters must be time-varying and in general have an impulse response corresponding to the maximum length signature observed ($p = k_{max}-1$). The Levinson recursion cannot be used to compute the classifier parameters. Further, design and application of a class-

ifier under nonstationary conditions generally assumes that the classifier will be applied starting at some exact time epoch. This is not possible in most practical situations.

Our purpose here is to develop a classifier structure for classes of signatures which can be modeled as containing a random trend. Such signatures frequently occur in practice and represent a form of temporal nonstationarity for which it is not necessary (indeed of no value) to know the time epoch. An example derived from observation of a satellite is shown in Fig. B.1(a). In spite of its random nature, the trend can be exploited for purposes of discrimination. In what follows, it will be shown that to account for the nonstationarity requires only a simple modification to the classifiers developed in Sections II and III.

It is easy to show that a linear trend in a set of (otherwise stationary) data can be removed by taking differences $\nabla \underline{x}_k = \underline{x}_k - \underline{x}_{k-1}$. Similarly any trend that can be represented by a polynomial of order d can be removed by taking differences of the data d times. Figure B.1(b) shows the data of Fig. B.1(a) after taking second differences. Except for small intervals at the nulls of the original signature the differenced data exhibits a stationary behavior. The resulting stationary process can be subjected to the usual processes of linear prediction in order to perform the discrimination.

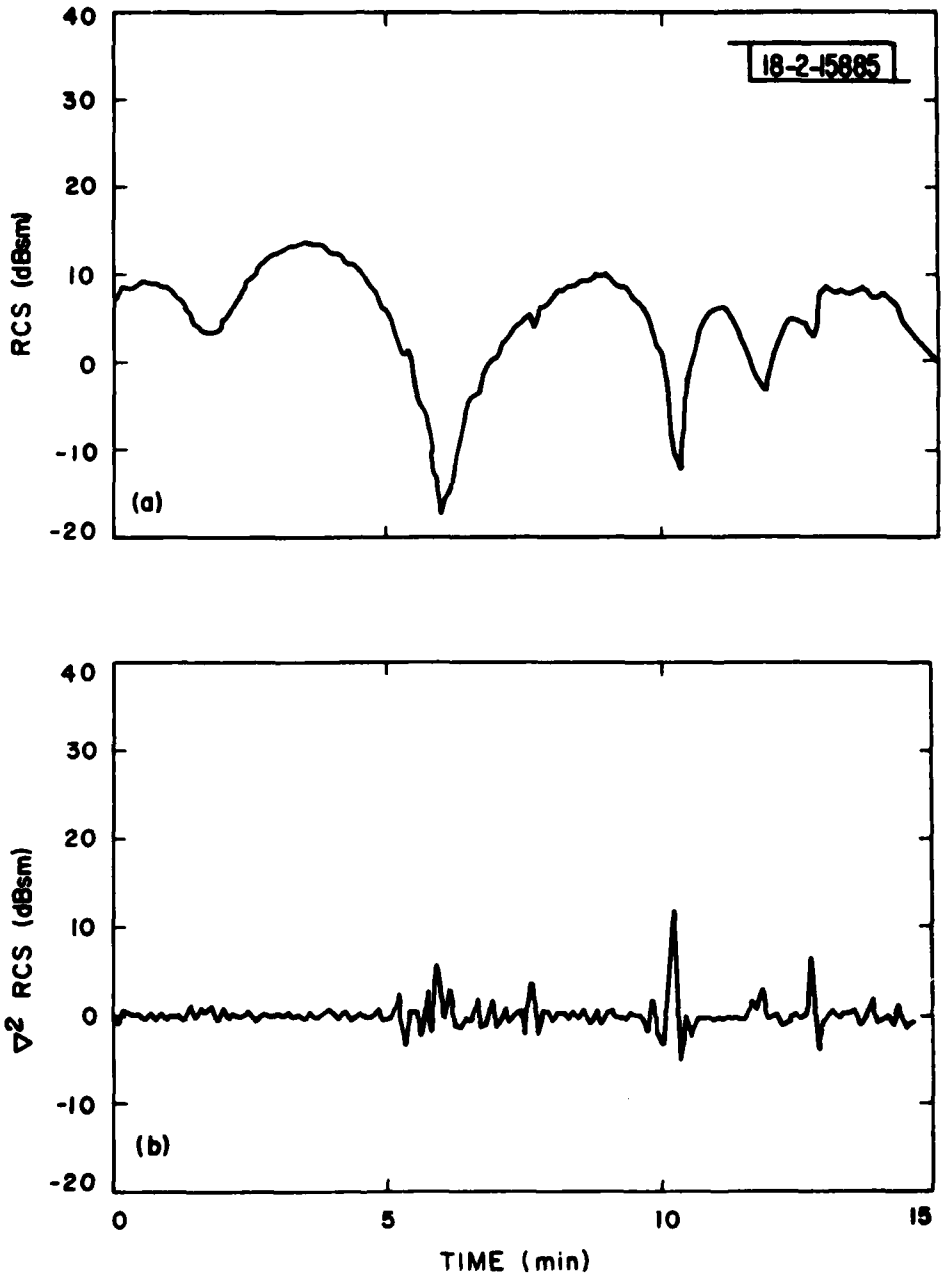


Fig.B1.(a-b) Removal of the trend in a signature by differencing.

To formalize the foregoing statements observe that if $\underline{u}_k = \underline{x}_k - \underline{x}_{k-1}$ then the set of differences $\underline{U}_k$ are related to the observations $\underline{X}_k$ by the linear transformation

$$\underline{U}_k = \begin{bmatrix} I & & & \\ -I & I & & \\ 0 & -I & I & O \\ & & & \\ 0 & \cdots & & -I \; I \end{bmatrix} \underline{X}_k \tag{B.2}$$

Since the Jacobian of this transformation is equal to one, we have

$$p_{\underline{X}}(\underline{X}_k) = p_{\underline{U}}(\underline{U}_k) \tag{B.3}$$

Thus if $\underline{U}_k$ is stationary we apply the arguments in the first part of this appendix to obtain the classifier structure. Otherwise higher order differences are taken until a stationary process is approximated and then the linear predictive structure is applied. In the case of a class of signatures with polynomial trend of order d, a branch in the classifier structure of Fig. 2.2 is replaced by the branch shown in Fig. B.2.

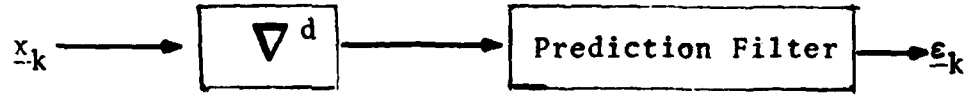


Fig. B.2 Branch of sequential classifier for nonstationary signatures.

The fundamental ideas in treating this class of non-stationary signatures are derived from the theory of ARIMA (autoregressive integrated moving average) models in time series analysis whereas the corresponding ideas for stationary models such as those discussed in the body of this report are derived from the theory of AR (autoregressive) models. For a thorough discussion of these topics see Ref. 15.

References

1. C. W. Therrien, "The Application of Linear Prediction to Sequential Classification of Radar Target Signatures," Technical Report 517, Lincoln Laboratory, M.I.T. (25 March 1976), DDC AD-A026695

2. C. W. Therrien, "A Sequential Approach to Target Discrimination," IEEE Trans, Aerospace Electron Systems, AES-14, 433 (1978)

3. F. C. Reed, "A Sequential Multidecision Procedure," Proc. Symp. on Decision Theory and Applied Electronic Equipment Development, Rome Air Development Center, Griffiss AFB, New York(April 1960), DDC AD-236338.

4. A. Wald, Sequential Analysis,(Wiley, New York, 1947)

5. K. S. Fu, Sequential Methods in Pattern Recognition and Machine Learning, (Academic Press, New York, 1968)

6. H. L. Van Trees, Detection, Estimation, and Modulation Theory, Part I (Wiley, New York, 1968)

7. M. G. Kendall and A. Stuart, The Advanced Theory of Statistics, Vol. 2, (Hafner Publishing Co., New York, 1973)

8. H. A. Akaike, "Time Series Analysis and Control through Parametric Models" in Applied Time Series Analysis (Academic Press, New York, 1978)

9. E. Parzen, "Multiple Time Series: Determining the order of Approximating Autoregressive Schemes," Multivariate Analysis IV, Proc. 4th Intl. Symp. 1975, P. R. Krishnaiah ed. (North Holland Publishing Co., New York, 1977)

10. A. Lindquist, "A New Algorithm for Optimal Filtering of Discrete-Time Stationary Processes," SIAM J. Control, 12, 736 (1974)

11. O. N. Strand, "Multichannel Complex Maximum Entropy (Autoregressive) Spectral Analysis," IEEE Trans. Automatic Control, AC-22 634 (1977)

12. A. Papoulis, Probability, Random Variables and Stochastic Processes (McGraw-Hill, New York, 1965)

13. H. L. Van Trees, Detection, Estimation, and Modulation Theory, Part III (Appendix) (Wiley, New York, 1971)

14. K. S. Miller, Complex Stochastic Processes (Addison-Wesley, Reading, MA., 1974)

15. G.E.P. Box and G. M. Jenkins, Time Series Analysis Forecasting and Control (Holden-Day, San Francisco, 1970)

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE *(When Data Entered)*

REPORT DOCUMENTATION PAGE	READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER: ESD-TR-79-308	2. GOVT ACCESSION NO.: AD-A084 272	3. RECIPIENT'S CATALOG NUMBER
4. TITLE *(and Subtitle)*: Application of Linear Prediction to Multiclass Sequential Discrimination	5. TYPE OF REPORT & PERIOD COVERED: Technical Note
	6. PERFORMING ORG. REPORT NUMBER: Technical Note 1979-63
7. AUTHOR(s): Charles W. Therrien	8. CONTRACT OR GRANT NUMBER(s): F19628-80-C-0002
9. PERFORMING ORGANIZATION NAME AND ADDRESS: Lincoln Laboratory, M.I.T. P.O. Box 73 Lexington, MA 02173	10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS: Program Element Nos. 63304A and 63308A
11. CONTROLLING OFFICE NAME AND ADDRESS: Ballistic Missile Defense Program Office Department of the Army 5001 Eisenhower Avenue Alexandria, VA 22333	12. REPORT DATE: 19 December 1979 13. NUMBER OF PAGES: 38
14. MONITORING AGENCY NAME & ADDRESS *(if different from Controlling Office)*: Electronic Systems Division Hanscom AFB Bedford, MA 01731	15. SECURITY CLASS. *(of this report)*: Unclassified 15a. DECLASSIFICATION DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT *(of this Report)*

Approved for public release; distribution unlimited.

17. DISTRIBUTION STATEMENT *(of the abstract entered in Block 20, if different from Report)*

TN-1979-63

18. SUPPLEMENTARY NOTES

None

19. KEY WORDS *(Continue on reverse side if necessary and identify by block number)*

linear prediction
sequential classifier
GSPRT
discrimination
target signature

20. ABSTRACT *(Continue on reverse side if necessary and identify by block number)*

A linear predictive form of classifier is described for the sequential discrimination of target signatures. The classifier implements a Generalized Sequential Probability Ratio Test (GSPRT) which allows for multiple classes of target signatures. Forms of the classifier for both noncoherent and coherent signatures are described. Flow charts are included for convenience of implementation.

DD FORM 1 JAN 73 1473 EDITION OF 1 NOV 65 IS OBSOLETE

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE *(When Data Entered)*